

# Research on Personalized Recommendation Algorithm for Internet User to Browse News

Chunsheng Li<sup>1</sup>, Shuchao Feng<sup>2</sup>, Wenqian Shang<sup>1</sup>

<sup>1</sup>School of Computer Science, Communication University of China, Beijing, 100024, China

<sup>2</sup>School of Science, Communication University of China, Beijing, 100024, China

**Keywords:** multi-Association rule, Fuzzy Theory, Membership Function.

**Abstract.** At present, the mode analysis methods for Internet User to browse news have three main kinds: that is, association analysis, clustering analysis and the mode of sequence analysis. This paper mainly focuses on the personalized news recommendation for Internet user. We combine the membership functions based on sample statistics distribution with the classical algorithms based on association rule. It can generalize its usefulness range. The structure of personalized news recommendation system is also given in this paper.

## Introduction

With the rapid development of Internet in recent years, personalized news recommendation has become an essential service for sites. Comparing with the booming e-commerce website, the service level of personalized news recommendation are still fall behind. An Internet user may not shop online, but most of them may read news online. In other words, there are more information websites users. So if the potential users' interest to better mining and giving corresponding personalized news recommendation, it can produce greater social and economic value. Hence, it's necessary to help Internet users find what they want to read automatically. This is called personalized news recommendation [1].

The format of the offline catching data often can be described as Table1:

Table 1: the section format of data

User-id	Press-id	Vis-time	Title	Content	Rel-time
100001	06227	2014/8/10:22	-----	-----	2014/7/30

From this Table, we can find that: the data includes user-id, press-id, visit-time, title, content and release-time. This means that we can use different algorithms from different views.

So far, there are many kinds of algorithms for personalized recommendation [2], which can be divided into four classed, that is, content-based recommendation algorithm, graph-based recommendation algorithm, recommendation algorithm based on collaborative filtering[3][4] and hybrid algorithm for personalized recommendation.

However, facing huge data, classical algorithms exits some disadvantages. For instance, the graph-based algorithm of personalized recommendation mainly constructs a bipartite graph based on user-content model. T. Zhou et al [5][6] construct a bipartite graph algorithm based on resource allocation matrix and random walk to realize personalized recommendation. But their computing and storage are most costly.

Through above analysis, this paper proposes an algorithm based on the theory of fuzzy mathematics. The algorithm can generalize the usefulness range of the classical. What's the most important is that the algorithm can mine some potential users who will not be discovered due to data cleaning error. Last but not least, we establish the model of personalized news recommendation for Internet users.

The rest of this paper is organized as follows: some definitions and theorems are introduced in next section. Section 3 presents an algorithm based on membership function. Section 4 presents the structure of personalized news recommendation and section 5 offers some concluding marks.

## Preliminaries

### Create membership Functions

The research on personalized news recommendation algorithm based on simulation dataset is from the third news portal. The distribution of the dataset can be described as Fig.1:

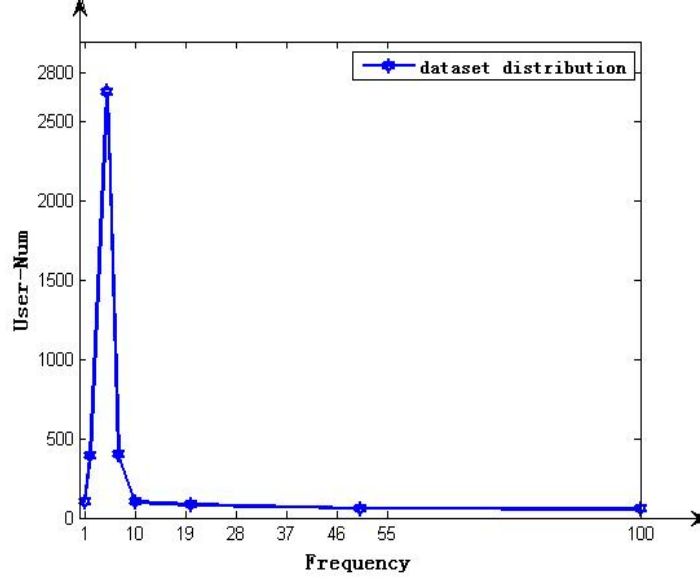


Fig.1: Users browsing frequency distribution

The Fig.1 shows the User-Num Frequency distribution for the dataset. From Fig.1, we can find that: the dataset can be divided into two parts, denoted by A and B, respectively.

A membership function defines the degree to which the value of a variable belongs to the group and is usually a linguistic such as high or low. We establish membership functions based on users habits of browsing news above referred. Their membership functions are as follows:

$$f_A(x) = \begin{cases} (x-2)/3, & 2 \leq x \leq 4 \\ 1, & 5 \leq x < 16 \\ \left[1 + \left(\frac{x-16}{2}\right)^2\right]^{-1}, & 16 \leq x < 80 \\ 0 & 80 \leq x \leq 1000 \end{cases} \quad (2.1)$$

$$f_B(x) = \begin{cases} 0, & 5 \leq x < 16 \\ (x-16)/78, & 16 \leq x < 55 \\ \left[1 + \left(\frac{x-50}{5}\right)^2\right]^{-1}, & 55 \leq x < 80 \\ 1 & 80 \leq x \leq 1000 \end{cases} \quad (2.2)$$

Where  $x$  is the number of items user has clicked. And  $f_A(x)$  is the degree to the group A.

### Maximum Membership Principles

#### Theorem 1

Suppose that  $X$  is a set of all objects to be identified, and  $A_i \in F(X)(i=1,2,\dots,n)$  represent  $n$  fuzzy patterns. For any element  $x$  in  $X$ , if the element  $x$  should be identified which model it belongs to, it can be judged by the following principle [7]:

$$\mu_A(x) = \max_{1 \leq i \leq n} \{\mu_{A_i}(x)\} \quad (3.1)$$

For example, we assume  $x=17$ , and then we have

$$f_A(17) = 0.8, \quad f_B(17) = 1/78$$

According to the maximum membership principle, the number 17 should belong to category attribute of A. Of course, the membership should be established by the distribution of the dataset.

### Pre-recommendation

Generally, the first step of data mining is to clear data, which will lose some information. In this section, we combine membership function with the theory of association rule to finish pre-recommendation for Internet Users to browse news. The pre-recommendation algorithm can be described as follows:

#### Pre-recommendation algorithm

Input: user-id, news-id and number  $x$ .

Output: recommendation lists of news-id with user-id.

Steps:

1. Construct membership functions.
2. For
3. Input  $x$  and choose maximum value, and  $x$  is deemed to category which is consistent with membership functions.

// In this step, the original data will be divided into two parts, A and B, respectively.

4. End for

5. If ( $x \in A$ ) then

Give min-support.

Scan data, record and save frequency of  $x$  presenting. // Generate 1-frequent Item.

Based on the step above, and generate 2-frequent Item.

Generate base rules if satisfying min-confidence.

6. Output recommendation result.

7. End.

In the algorithm above, there are several notes needed to be noticed. Firstly, the data we meet may not be like the table.1. This means that we need to create membership functions in accordance with specific data. With that we can generate category attribute respectively.

The flow diagram of pre-recommendation can be described as Fig.2:

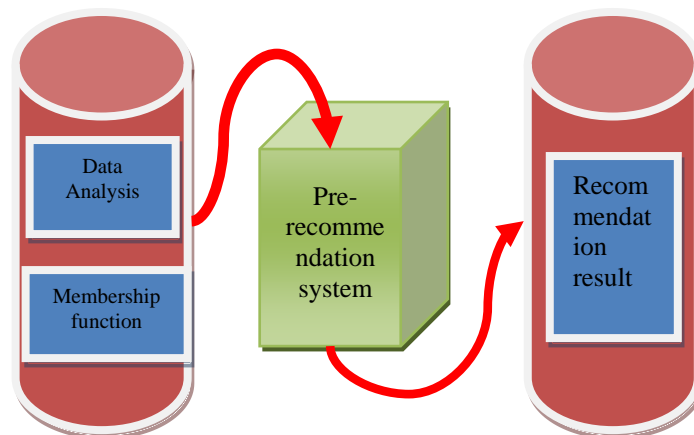


Fig.2: The flow diagram of pre-recommendation

### Optimization Algorithm of Personalized Recommendation

Internet Users have got recommendation result in section 3, but the data is rough after all. We need further optimize the algorithm of personalized news recommendation. The flow diagram of generating recommendation result can be described as Fig.2:

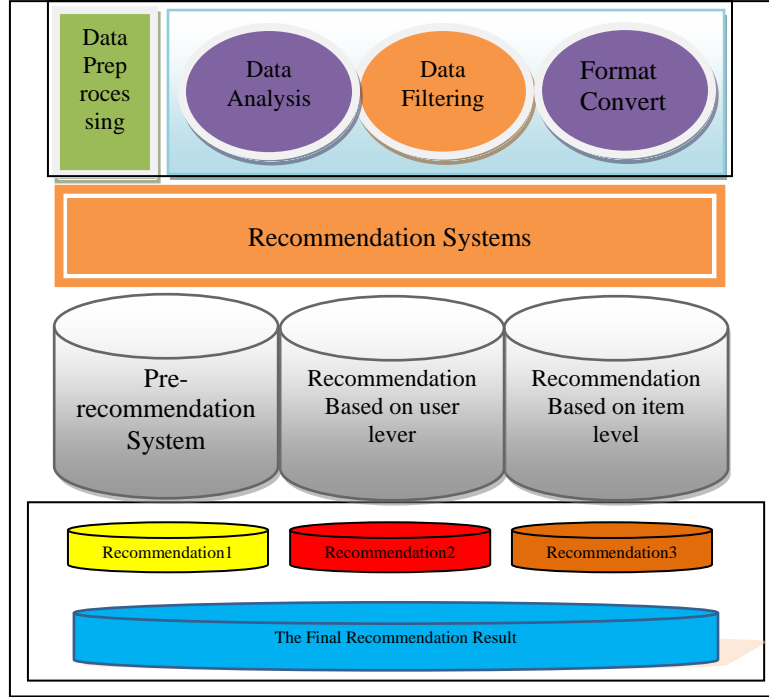


Fig.2: the flow diagram of generating recommendation

S. Z. Zhang et al [8] present a hybrid graph model with two layers for personalized recommendation based on small world network and Bayesian network. Assuming that there are 3 users and 5 news documents in recommendation system, the system can be express by eight node bipartite graph. It can be described as Fig.3:

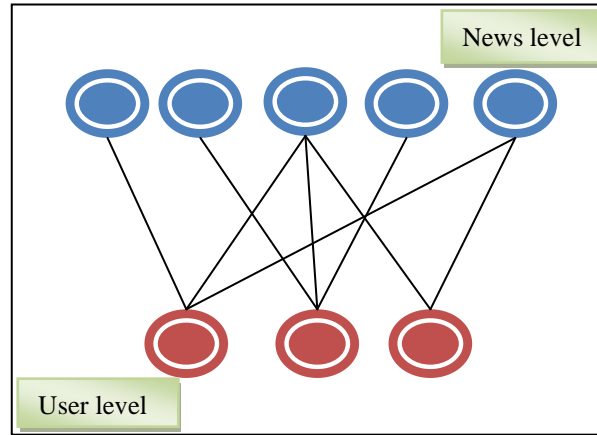


Fig.3: A Graph based on User- News

This paper also constructs the model of recommendation algorithm for Internet User browsing news. The algorithm includes two parts. One is user lever, the other is item level. The algorithm based on user level is as follow:

Step1: Generate the vector model. The element of vector is item which user has clicked.

Step2: Calculate similarity. The similarity between two users arbitrarily is measured by Jaccard distance.

$$J(u_i, u_j) = \frac{|C_{u_i} \cap C_{u_j}|}{|C_{u_i} \cup C_{u_j}|} \quad (4.1)$$

Where  $|C_{u_i} \cap C_{u_j}|$  represents the number of the same items user  $u_i$  and  $u_j$  have clicked.  $|C_{u_i} \cup C_{u_j}|$  is the total number of the different items user  $u_i$  and  $u_j$  have clicked.

Step3: Generate the recommendation result. Recommending interest items for users can find other users who have similar preference, which is the basic assumption of recommendation

algorithm based on collaborative filtering.

Step4: End.

The algorithm based on item level is as follow:

Step1: Pre-processing the text documents and generating the text vector space.

While we must deal with the format of the original data before generating vector space. The text document pre-processing often includes words segment, extracting stems, removing stop words and selecting features [9].

Step2: Calculate similarity. The measure between text  $d_1$  and text  $d_2$  is named similarity. For any two news  $d_1(x_{11}, x_{12}, \dots, x_{1n})$  and  $d_2(x_{21}, x_{22}, \dots, x_{2n})$ , we can define it by cosine formula.

$$\text{sim}(d_1, d_2) = \cos \theta = \frac{\sum_{i=1}^n x_{1i} \cdot x_{2i}}{\sqrt{\sum_{i=1}^n x_{1i}^2 \cdot \sum_{i=1}^n x_{2i}^2}} \quad (4.2)$$

Choose the max-similarity item and recommend it for user.

Step3: Consider the sequence of time.

Considering that the recommendation result has been got, we should make a decision that users can browse the recommendation items. This means that the visiting time must be after the releasing time.

Step4: Optimize recommendation result.

The step means that we need to make last decision for recommendation result.

Step5: End.

In fact, the methods of optimizing the algorithm about personalized recommendation of Internet User browsing news are various, including other classical algorithms, like Naive Bayes decision, Bayes Belief Network, and kinds of clustering ways. Of course, there are many academic achievements that improved the original algorithms. For instance, Y.W.Qiao and C.J.Chen provide new Apriori algorithm based on automatic fuzzy partition [10].

## Conclusions

According to the principle of membership function and the maximum membership, the uncertainty that is produced by human factors can be described better. Based on fuzzy theory, classical association rules can be applied to bigger dataset. This paper gives the process of establishing a model of personalized news recommendation. In the future, we will further study the personalized news recommendation algorithm.

## Acknowledgements

This paper is partly supported by the Project of Comprehensive Reform of the Major of Computer Science and Technology (Project No. ZL140103) and partly supported by Engineering Project Planning of Communication University of China (Project No. XNG1436).

## References

- [1] R.Q.Xie. Personalized news recommendation via implicit social experts [D]. Xiamen University. 2014(5).
- [2] Yang Bo, Zhao pengfei. Recommendation algorithm overview[J], Journal of Shanxi University(Nat.Sci.Ed),2011,34(3):337-350.
- [3] AN Optimized Item-Based Collaborative Filtering Recommendation Algorithm[A].2009 - 2009 IEEE International Conference on Network Infrastructure and Digital Content.2009.
- [4] Xiaoyuan Su, Taghi M Khoshfotaar. A survey of collaborative filtering techniques[J]. Advanced in Artificial Intelligence,2009,2009(4).

- [5] Zhou T, Ren J, Medo M, et al. Bipartite network project and personal recommendation. *Phys Rev E*, 2007,76:046115
- [6] Zhou T, Jiang LL, Su RQ, et al. Effect of initial configuration on network-based recommendation. *Europhys Lett*, 2008, 81:58004.
- [7] BP Neural Network Algorithm for Multi-sensor Trajectory Separation Based on maximum Membership Degree [A]. 2013 3rd International Conference on Computational Intelligence and Industrial Application. 2013.
- [8] Zhang Shaozhong, Chen Deren. Hybrid graph Model with two layers for personalized recommendation[J]. *Journal of Software*, 2009(12)123-130.
- [9] W.Q. Shang. Research on Text Categorization [D]. Beijing: Beijing Jiaotong University, 2007.
- [10] Y.W.Qiao,C.J.Xiao,H.Yang. Generation of QAR Association Using Rules Based on Automatic Fuzzy Partition and Improved Apriori Algorithm [J]. *Electronics Optics & Control*.2012(5)36-40.
- [11] J.J. Xie, C.P. Liu. The method of fuzzy mathematics and its application [M]. Wuhan: HuaZhong University of Science and Technology Press,2013.2.